NASA 2003 Faculty Fellowship Program Lecture Series:

"*Unleashing Error or Uncertainty Analysis of Measurement Results*"

By

M. Kostic

Department of Mechanical Engineering

Northern Illinois University

www.kostic.niu.edu

Uncertainty or error analysis of measurement results is not a deterministic (exact), but rather holistic and probabilistic in nature. Its complexity and ambiguity (*not only what is measured, but what contributes to measurement errors and uncertainty, the latter being open ended*), contributes that the uncertainty analysis is often misunderstood, misrepresented or even avoided. An almost impossible task is attempted here: to try to resolve existing confusion, to embrace the very concept of measurement uncertainty, and to provide effective guidelines to account for the most contributing sources of errors (which is important and possible!), since accounting for all sources of measurement errors is not necessary (and also impossible).

About the Presenter:

M. KOSTIC is an Associate Professor in the Department of Mechanical Engineering at Northern Illinois University. He received his Ph.D. from the University of Illinois, and then worked in industry for some time. Professor Kostic's teaching and research interests are Thermodynamics, Fluid Mechanics, Heat Transfer and related energy sciences; with emphases on new technologies, experimental methods, creativity, design, and computer applications (www.kostic.niu.edu).

At **Ohio Aerospace Institute** Lecture Hall

Friday, August 1, 2003, 2:00 p.m.,

Refreshments Available at 1:30 p.m.

Presentation handouts:

http://www.kostic.niu.edu/NASA-FFP03Kostic(D).pdf

Manuscript:

http://www.kostic.niu.edu/Uncertainty-Analysis-of-Measurement-Results.pdf

References

[1] Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results, NIST Technical Note 1297, 1994 Edition http://physics.nist.gov/Pubs/guidelines/TN1297/tn1297s.pdf , July 2003.

[2] Galleries of Probability Distributions: *NIST/SEMATECH e-Handbook of Statistical Methods*, http://www.itl.nist.gov/div898/handbook/eda/section3/eda366.htm , July 2003.

[3] R.H. Dieck, "Measurement Accuracy" a chapter in "The Measurement, Instrumentation and Sensors Handbook" *(J.G. Webster, Editor-in-Chief)*, ISBN: 0-8493-8347-1, CRC Press, 1999.

[4] *NIST/SEMATECH e-Handbook of Statistical Methods*, http://www.itl.nist.gov/div898/handbook/, July 2003.

[5] Tables for Probability Distributions: *NIST/SEMATECH e-Handbook of Statistical Methods*, http://www.itl.nist.gov/div898/handbook/eda/section3/eda367.htm , July 2003.

[6] R.S. Figliola, and D.E. Beasley, *Theory and Design for Mechanical Measurements* - 3rd Edition, Wiley, 2000.

→

Error or Uncertainty Analysis of Measurement Results

by M. Kostic , < www.kostic.niu.edu > kostic@niu.edu
@ NASA Glenn Research Center < www.grc.nasa.gov > Milivoje.Kostic@grc.nasa.gov
Summer 2003

1. Concept and Definitions

Uncertainty or error analysis of measurement results is not a deterministic (exact), but rather holistic and probabilistic in nature. Its complexity and ambiguity (not only what is measured, but what contributes to measurement errors and uncertainty, the latter being open ended), contributes that the uncertainty analysis is often misunderstood, misrepresented or even avoided. An almost impossible task is attempted here: to try to resolve existing confusion, to embrace the very concept of measurement uncertainty, and to provide effective guidelines to account for the most contributing sources of errors (which is important and possible!), since accounting for all sources of measurement errors is not necessary (and also impossible).

A measurement outcome or result is only reliable (i.e. meaningful) if its accuracy or uncertainty of relevant errors is quantified, since it is only an approximation (or estimate) of the specific quantity subject to the measurement. Only if the uncertainty of a measurement result is quantified, it may be judged if the result is adequate for intended purpose, i.e. useful, or consistent with other similar results.

However, all measurements are intrinsically predisposed to different sources of errors and, thus, it is impossible to measure or otherwise obtain a perfect result without any error. Therefore, it is only possible to reasonably estimate the actual (or true) measured quantity within a specified range or interval of possible errors (called *uncertainty*), and even that is not absolutely certain, but only probable within a specified *confidence level* (for an arbitrary, say 95% probability). However, it is factually experienced that the results of repeated measurements of apparently the same quantity (called *measurand*) under apparently the same (or very similar) controlled conditions, are not exactly the same but tends to conglomerate around a certain value (*modal value*, equal to the *mean value* for symmetric distribution).

The objective of uncertainty analysis is to scientifically estimate the uncertainty range of possible measurand error from reported measured result within a specified probability, or vice versa, to estimate probability for a reported uncertainty, since the two are uniquely correlated to each other for a given conditions. It is important to clarify again, that only the uncertainty, i.e. a deviation range (or interval) from a reported measurement result, with corresponding probability, may be evaluated, but it is not possible to obtain a perfect (error-free) measurement, nor it is possible to estimate an uncertainty with 100% probability (absolute certainty is impossible too). However, under well-controlled conditions and well-understood measurement process and procedure, it is possible to minimize and reasonably well (at high probability) estimate the uncertainties of measured quantities and the final measurement result.

Evaluation (calculation or quantification) of uncertainties is very judgmental process, based on full scientific understanding, knowledge and experience of measured quantities, properties and characteristics of instrumentation used, measurement processes and procedures, and the data reduction procedures used to obtain the final measurement result. It is impossible to account for all sources of errors, but it is critical (and possible!) to account for the most important ones (i.e. the most contributing ones, which is also satisfactory). Uncertainties could be classified in two categories or types [1]: The first *type A uncertainties* are due to random errors (also known as precision, or statistical uncertainties), which could be quantified using available statistical methods on a set of repeated measurements under apparently the same (or very similar) conditions. Additional uncertainties classification, *type B uncertainties,* are due to other than statistical errors associated to instrumentation and measurement procedure used (often unspecified, also known as biased or systematic errors). These, type B uncertainties, are evaluated using scientific judgment and previous related experience and calibration, including instrumentation specifications, other outside sources and references, and judgmental assumptions. Distinction between the two types of uncertainty (statistical type A and non-statistical type B) is not always clear, and meaning of many terms, like systematic error for example, may be misleading and should not be used exclusively, but only in relative terms. Systematic errors and specified biases if quantified (like after careful calibration) are no longer uncertainties, since they may be corrected, but only unspecified biases (due to unknown/random nature) are inherited uncertainties of instrumentation and similar, previous measurements. For example, a biased instrument uncertainty (when an instrument is used) is a type B uncertainty, based on previous calibration of that or similar instruments and usually obtained from outside sources; however, it was in part the type A statistical uncertainty during calibration and testing of that or similar instruments.

2. Evaluation/Quantification of Uncertainties

Even though uncertainty quantification is very judgmental and based on full physical understanding of all related issues and accumulated previous experience (it is more an art than a science), the uncertainty analysis procedure is rather simple, and consists of the following steps (see also an example below):

(1) identification of all critical (i.e. the most contributing) uncertainty components (this is very important since the final result depends on what is accounted for);
(2) quantification of all standard uncertainties, based on relevant statistics, like the *Student's t-distribution* to account for small number of measurements (type A uncertainties), or obtained from other external sources (type B uncertainties);
(3) use the law of propagation of uncertainties (sensitivity coefficients based on the first-order Taylor series) when a partial or final measurement result is determined using an equation from other measured quantities;
(4) reduce all uncertainty components to standard uncertainties or to uncertainties for the same probability confidence level (opposite from expended uncertainty, see below) and combine them using the Root-Sum-of-Squares (RSS) rule;
(5) Expend the combined uncertainty to desired confidence level by multiplying it with a relevant coverage factor or dimensionless deviation t, based on the combined effective degree of freedom using the Student t-distribution.
(6) Finally, report all relevant details of accounted uncertainty components along with expended resultant uncertainty, but also some indication of the values of uncertainty components not accounted for, should always be reported.

The above steps will be detailed below. However, the final uncertainty results, regardless if the same analysis procedure is used on the same measured data, depends on the above step (1) and step (2) type B uncertainties, both being very judgmental and critical for the final results. Thus, the uncertainty of the result of a measurement is an estimate and can not have well-defined limits. Therefore, only a range or interval of possible errors may be quantified within some *confidence level* (*probability*), called measurement *uncertainty*, while absolute error (difference between measured value and true measurand) cannot be quantified and is a qualitative concept only.

However, the precision or random error (difference between a measured and their mean values) can be quantified. Many other terms, like precision, systematic error and bias, must be accurately defined and are often misleading. Fundamentally, statistics is based on large number of data, with an assumption that the uncertainty of the standard deviation is negligible. However, for a small number of measurements (often the case), not every estimated standard deviation is necessarily a standard uncertainty. In general, the measurement uncertainty depends not only on the *repeatability* (under the same measurement conditions) and *reproducibility* (under different conditions), but also on how well a measurement method has been chosen and implemented. As stated above, some indication of the values of uncertainty components not accounted for, should always be reported. Again, not only a measurement result, but also its uncertainty, are both uncertain within some confidence level.

2.1. Type A, Statistical (Precision) Uncertainties

A measurement uncertainty is a probable range of measurement errors, it is intrinsically random in nature (since measurements can not be perfectly controlled and executed) and it could be well-approximated with some kind of statistical distribution. If no specific physical reason dictates a special statistical distribution [2], most of the random errors could be well-approximated with standard, *normal Gaussian distribution* for large number of measurements, or with the corresponding *Student's t-distribution* for small number of measurements. The measurand, a specific quantity X, subject to set of n measurements x_i, is best represented by the mean value, x_m, and a measure of random scatter of data sample may be represented by the *sample standard deviation*, S_x, using the well-known correlations:

$$x_m = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad (1)$$

$$S_x = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\underbrace{(x_i - x_m)}_{d_i}{}^2} = \sqrt{\frac{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}{n(n-1)}} = \sqrt{\frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - n x_m^2\right)} \qquad (2)$$

Similarly, the standard deviation of data $\{x_i, y_i(x_i)\}$, curve-fitted with a function $y_{cf,i}=y_{cf}(x_i)$, is calculated using the following correlation:

$$S_{xy} = \sqrt{\frac{1}{\nu_{cf}}\sum_{i=1}^{n}(y_i - y_{cf,i})^2}\,;\quad \nu_{cf} = n - (\text{no. of unknown coeffs. in } y_{cf}) \qquad (3)$$

Note that Eq. (2) is a special case of Eq.(3) if $y_i = x_i$, $y_{cf}(x_i) = x_m$ and $\nu_{cf} = n-1$.

The *mean-value standard deviation,* S_{xm} , is becoming smaller with increasing the number of measurements, n (thus n could be adjusted to achieve a desired mean-value uncertainty), and is expressed as:

$$S_{xm} = \frac{S_x}{\sqrt{n}} = \sqrt{\frac{1}{n(n-1)}\sum_{i=1}^{n}(x_i - x_m)^2} \qquad (4)$$

Assuming that a measurement uncertainty follows the normal Gaussian distribution (often very reasonable if measured data are bell-shaped and symmetric), it is expected that about 68.27% of all measurements (if the number is large enough) should be within $\pm S_x$ deviations from the mean value, while the remaining 31.73% are expected to fall outside the $\pm S_x$ deviations from the mean value. The uncertainty probability (or confidence level) will increase to about 95.45% for $\pm 2S_x$ deviations from the mean value, and to 99.73% for $\pm 3S_x$ deviations from the mean value, according to the Gaussian distribution. The probability is uniquely related to the deviation range (i.e., uncertainty) for a given statistical distribution and vice versa, and may be determined from the corresponding tables or calculated using related functions in many standard application software, like Excel, MathCAD, MATLAB, etc., see also Table 1 and Refs. [1], [3], and [4].

If another set of measurements is repeated, with the same number of measurements, n, and under similar conditions as before, it is expected that the new mean value will be close to the previous mean value, actually that about 68.27% of all such mean values (if the number is large enough) will be within $\pm S_{xm}$ deviations from the mean of the mean values (overall mean), while the remaining 31.73% will fall outside the $\pm S_{xm}$ deviations from the overall mean value (remember that $S_{xm} = S_x / \sqrt{n} < S_x$). The uncertainty probability will increase to about 95.45% and 99.73% for $\pm 2S_{xm}$ and $\pm 3S_{xm}$ deviations from the overall mean value, respectively.

If number of measurements is not large enough (say $n<20$ or so, which is often the case), it is more accurate to use the Student's t-distribution instead of the normal Gaussian distribution. The concept and procedure is the same except that values of probability are dependent, in addition to the deviation range, also on degree of freedom (ν), which in turn is dependent on number of measurements (for a measured constant value the $\nu=n-1$). For the same deviation, for example, for $\pm S_x$ deviations from the mean value, the probability of the Student's t-distribution will increase with number of measurements from 50% for $n=2$ (i.e. $\nu=1$) to 65.66% for $n=10$, and will approach the Gaussian's 68.27% probability for a large number of measurements. The Student's t-distribution tables and functions are available along with normal Gaussian and other statistical functions in many standard references [5] and software. However, the relevant data are usually presented in dimensionless forms (to avoid any issues related with units), and dimensionless deviations or intervals (sometimes referred to as *coverage factors*) are defined as:

$$t_k = z_k = \frac{d_k}{S_k} = \frac{x_k - x_{km}}{S_k} \quad (\text{NOTE } d_{k,i} = x_{k,i} - x_{km}\text{; } d_k \text{ is desired limit of } d_{k,i's}) \tag{5}$$

The statistical parameters are mutually and uniquely correlated for different statistical distributions, for example:

$$Student\_t\_distribution(t, \nu, P\%) = 0 \tag{6}$$

$$Gaussian\_distribution(z, P\%) = 0\text{ ; } (\nu \to \infty \Rightarrow t \to z) \tag{7}$$

where t (or z, see Eq. 5) is the Student t-distribution interval, a statistical function of degree of freedom ν, and confidence level $P\%$, which uniquely correlates the three quantities (Eq. 6), and it reduces to the normal Gaussian correlation for infinite (or large enough) degree of freedom (Eq. 7), see also Figure 1. From implicit correlations, Eqs. (6 or 7), any parameter may be explicitly expressed as a corresponding statistical function of the remaining parameters, available in standard software or tabulated in references. Different symbols, z_k and t_k, for otherwise same quantity (and concept), namely dimensionless deviation, also called the *coverage factor,* are used to indicate if the Gaussian or Student's t-distributions are used, respectively. As stated before, the probability for a given statistical distribution and number of measurements (i.e. degree of freedom) is dependent on the deviation magnitude only and vice versa. Both the Gaussian and Student's t-distributions are symmetrical around the mean value (or zero dimensionless deviation) and integral over each side is 50%, to make up the total probability of 100% for all deviations possible (when the dimensionless deviation ranges from negative to positive infinity). In different references (and software) different portions (fractions) of probabilities are given for the same dimensionless deviation which may introduce unnecessary confusion. However, those different portions of probabilities are simply correlated and thus conversion is straightforward when required, see Figure 1 and Table 1. Either an inside probability, ½P_{in}, is given for positive deviation range, or ±double inside, symmetrical probability, P_{in}, is given. Also corresponding outside, single-tailed probability, ½P_{out}=50%-½P_{in}, or double-tailed outside probability, P_{out}=100% -P_{in}, may be given. Furthermore, so called quantile or cumulative probability for a deviation range from negative infinity to an arbitrary value is given, for example in MathCAD, and it is related to one-sided inside probability as: $P_z=P_{upto\_z}=0.5+sign(z)\ ½P_{in|z|}$. Since Student's *t*-distribution accounts for (small) number of measurements and reduces to normal Gaussian distribution for large number of measurements, it is recommended to be used always, because it accounts for the uncertainty of the standard deviation itself, the latter assumed to be negligible for large number of measurements when the Gaussian distribution may be used.

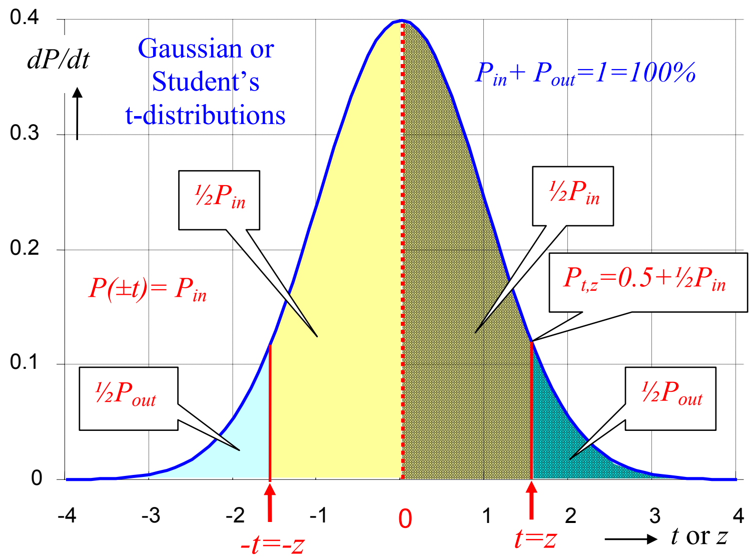

Figure 1: **Relation between uncertainty (t or z) and corresponding probability (P)**

Given, in Table 1 below, are the corresponding Excel functions and formulas along with nomenclature introduced here, to calculate important statistical parameters, some of them available in Excel's Descriptive Statistics in Data Analysis submenu within the Tool menu (if add-on Analysis *ToolPak* is installed, see the last column in Table 1).
Therefore, for a desired probability, *P%*, the dimensionless deviation, *t*, is evaluated first, using relevant tables or software (like in row 16 in Table 1 below), and then the dimensional deviation or uncertainty is calculated using the corresponding sample- (or mean-, if appropriate) standard deviation, i.e.:

$$d_{\nu,P\%} = u_{\nu,P\%} = t_{\nu,P\%} \cdot S_{x(m)} \quad \textbf{(8)}$$

Conversely, for a given or desired dimensional deviation *d* (i.e. uncertainty *u*), and degree of freedom, *ν=n-1*, the dimensionless deviation, *t*, is evaluated first, using the corresponding sample (or mean-value standard deviation if appropriate), i.e.:

$$t_{\nu,P\%} = \frac{d_{\nu,P\%}}{S_{x(m)}} = \frac{u_{\nu,P\%}}{S_{x(m)}} \quad \textbf{(9)}$$

and then the corresponding (double-sided-inside) probability, $P_{\nu,t}$, is evaluated using relevant tables or software (like in row 17 in Table 1 below).

It is straightforward to recalculate a known (old) deviation or uncertainty at one (old) given probability to a new deviation or uncertainty for another (new) desired probability or vice versa, as described in Section 2.5.

Table 1: Statistical parameters and related Excel functions

Row #	*Current nomenclature*	Excel functions and formulas	Excel Descriptive Statistics' items
1	x_m	AVERAGE(Data)	Mean (average)
2	S_{xm}	STDEV(Data)/SQRT(COUNT(Data))	Standard Error (of mean)
3		MEDIAN(Data)	Median (mid-way of data)
4		MODE(Data)	Mode (most frequent data)
5	S_x	STDEV(Data)	Standard Deviation (of data sample)
6	S_x^2	VAR(Data)	Sample Variance
7		KURT(Data)	Kurtosis (peakedness)
8		SKEW(Data)	Skewness (asymmetry)
9		MAX(Data)-MIN(Data)	Range
10		MIN(Data)	Minimum
11		MAX(Data)	Maximum
12		SUM(Data)	Sum
13	n	COUNT(Data)	Count (number of data)
14		LARGE(Data, i_th)	Largest(i_th=1)
15		SMALL(Data, i_th)	Smallest(i_th=1)
16	$u_{v,P\%}=t_{v,P\%}\cdot S_{xm}$	TINV(1-Probability, COUNT-1)*Standard_Error	Uncertainty at Confidence Level (for 95.0%=Probability)
17	$P_{t,v}$	1-TDIST(t, COUNT-1,2)	Probability (Double-Inside) for given deviation *t*

2.2. Type B, Non-Statistical (Bias) Uncertainties

As already stated, in addition to uncertainties determined using relevant statistical methods for measured data (subject to measurement repeatability and reproducibility), also the uncertainty associated with other inherited instruments and measurement procedure errors have to be evaluated and combined into a resultant uncertainty at a desired level of confidence. These other uncertainties are inherited (from all time history) for the instruments and measurement methods chosen and used at a time. They are estimated from reliable external reference sources or experimenters' previous experience (personal or others' accumulated knowledge about the instruments and methods). All uncertainties (including so-called non-statistical) are intrinsically random and statistical in nature and may be quantified to some confidence level (or probability) only. Even, so-called biased errors are really random and statistical, since we are not sure whether they are positive or negative, and the magnitude range is only estimated within a specified confidence, based on previous statistical observations of many similar instruments and measurement methods. Once, a so called bias error is systematically verified and specific correction is determined, it is no longer uncertain, thus that part of uncertainty is removed and reduced to the remaining random uncertainty (up to the random error of an instrument and testing method). Thus, a bias, non-statistical instrument uncertainty, represents a possible error range within a specified confidence level for a group of similar instruments manufactured and tested to be used within a certain specifications. Those so-called bias or non-statistical uncertainties actually were random and statistical uncertainties during testing and calibration of these instruments. Thus, even the new term, non-statistical (type B) uncertainty [1], is as misleading as old bias or systematic uncertainties. Intrinsically, the errors and uncertainties are random and statistical, and only distinction is whether the random scatter is observed and processed by an experimenter at present time (type A) or inherited from others' long time experiences (type B uncertainty, passed to us from previously experienced and accumulated type A uncertainties by others). The instruments' type B uncertainties are usually given by manufacturers or estimated and given in references for a well-known class of instruments manufactured within standard specifications. The most critical is estimation of type B measurement methods' uncertainties, since they depend on the employed methods of measurements and care of relevant details. For example, if we are manually timing a process with a stop-watch of 1 millisecond of instrument uncertainty, but starting and stopping synchronization process is uncertain within several tenths of a second, then the latter uncertainty is more contributing to the result than the former one. That why the knowledge and experience is more critical for proper quantification of the type B than type A uncertainties, since the former ones are less obvious. Generally speaking, the type B uncertainty quantification depends on the available external information and the experience and assumption of an experimenter.

2.3. Sensitivity or Dependence of Result on Other Quantities: The Law of Propagation of Uncertainties

If a measured result, x_R, is a function of (i.e. depends on) other measured quantities, x_i, with respective uncertainties u_i,

$$x_R = f(x_1, x_2 ... x_i ..., x_N) \tag{10}$$

then, there will be an uncertainty in result due to each of uncertainties of measured quantities x_i, i.e.:

$$u_{R,i} = \left|\frac{\partial f}{\partial x_i}\right| u_i = c_i u_i ; \quad (1 < i < n) \tag{11}$$

The above is a partial uncertainty in result $u_{R,i}$, due to dependence on a measured quantity x_i, and its uncertainty u_i. The dependence or sensitivity of the measured result on the measured quantity x_i, is expressed by magnitude of the corresponding partial derivative (based on the first-order Taylor series coefficients) and is also known as *sensitivity* or *dependency coefficient*, c_i. Regardless of the different dimensions of the measured quantities x_i, and their uncertainties u_i, after multiplication with the corresponding partial derivatives (sensitivity coefficients) they will all result in the same dimension of the resultant quantity and could be combined together (see Section below). If a particular partial derivative has a zero value, it means that the result is not dependent on that particular measured quantity, the latter uncertainty will not propagate and affect the resultant uncertainty. Conversely, if the magnitude of the partial derivative (thus dependence or sensitivity) is very large, then that component uncertainty will propagate very strongly (will amplify) and may adversely influence the resultant uncertainty.

2.4. Combining Uncertainties Due to Different Sources: RSS Method

Since there are many different sources of uncertainties for the same measured (x_k) and/or resultant quantity (x_R), they will increase the total uncertainty of that quantity. It is unlikely for many uncertainties (due to their random nature) to have the same sign, so it would be inappropriate to combine them by adding their magnitudes, since many of them will be in opposite directions and cancel each other to some extent. It is more probable to combine uncertainties of the same quantity (say x_R) into a resultant or total, or *combined uncertainty, $u_{R,c}$*, using the so-called Root-Sum-of-the-Squares (RSS) rules:

$$u_c = u_{R,c} = \sqrt{u_{R,1}{}^2 + u_{R,2}{}^2 + ... + u_{R,N}{}^2} = \sqrt{\sum_{i=1}^{N} u_{R,i}{}^2} = \sqrt{\sum_{i=1}^{N}\left(\left|\frac{\partial f}{\partial x_i}\right| u_i\right)^2} = \sqrt{\sum_{i=1}^{N}(c_i u_i)^2} \quad (12)$$

and the corresponding combined degree-of-freedom is [6]:

$$\nu_c = \nu_{R,c} = \frac{\left(\sum_{i=1}^{N}(c_i u_i)^2\right)^2}{\sum_{i=1}^{N}\frac{(c_i u_i)^4}{\nu_i}}$$

Note that ν_i (or ν_c) for type B uncertainty is usually very large (all time history) and may be taken as infinity if not specified otherwise, i.e. the Gaussian distribution may be used. It is important that every component of uncertainty, $u_{R,i}$, due to different source of uncertainty be mutually uncorrelated (zero covariance), be at the same confidence level (probability), and of course they all should have the same dimension expressed in the same units of measurement. Every uncertainty component may be estimated independently (thus, $c_i=1$), or determined using the law of propagation of uncertainties, Eq. (11), if the source of uncertainty is due to sensitivity (dependence) on measured component quantity ($c_i=|\partial f/\partial x_i|$).

2.5. Expended Uncertainty to Desired Confidence Level

It is straightforward to recalculate a known (old) deviation interval or uncertainty at one (old) probability to a new deviation or uncertainty for another (new) probability, called *"expanded uncertainty,"* i.e.:

$$u_{newP\%} = \frac{t_{\nu,newP\%}}{t_{\nu,oldP\%}} \cdot u_{oldP\%} \quad \text{or} \quad d_{newP\%} = \frac{t_{\nu,newP\%}}{t_{\nu,oldP\%}} \cdot d_{oldP\%} \quad \textbf{(13)}$$

where $t_{\nu,newP\%}$ is the Student *t*-distribution interval (also called the *coverage factor*), a statistical function of degree of freedom ν (or ν_c if appropriate), and confidence level *P%*, which uniquely correlates the three quantities and reduces to the normal Gaussian correlation for infinite (or large enough) degree of freedom, see Eqs. (6 & 7). This conversion will be needed for all types of uncertainties (if available at different probability confidence level) to bring them to the same probability level, before combining them using the root-sum-of-squares (RSS) procedure, Eq. (12), described above. If $d_{oldP\%}$ (or $u_{oldP\%}$) = $S_{x(m)}$ then $t_{\nu,oldP\%}=z_{oldP\%}=1$, i.e.:

$$u_{newP\%} = d_{newP\%} = t_{\nu,newP\%} \cdot S_{x(m)} \quad (14)$$

Also, for any desired deviation interval, $d_{newP\%}$ (or $u_{newP\%}$), the corresponding confidence level (i.e. probability), *P%*, may be determined using relevant statistical tables or formulas. It will be necessary to first determine the corresponding dimensionless deviation, *t* or *z*, based on given standard deviation, $S_{x(m)}$, or on a known uncertainty interval at a known confidence level, $u_{oldP\%}$, i.e.:

$$t_{newP\%} = \frac{u_{newP\%}}{S_{x(m)}} = \frac{d_{newP\%}}{S_{x(m)}} = z_{newP\%} \quad (15)$$

$$t_{\nu,newP\%} = \frac{u_{newP\%}}{u_{oldP\%}} t_{\nu,oldP\%} \Big|_{\nu\to\infty} = \frac{u_{newP\%}}{u_{oldP\%}} z_{oldP\%} = z_{newP\%} \tag{16}$$

Equation (15) is a special case of Eq. (16) when $u_{oldP\%}=S_{x(m)}$ since then $t_{\nu,oldP\%}=z_{oldP\%}=1$. Conversely, for a desired confidence level *P%*, the corresponding dimensionless deviation or uncertainty may be determined using relevant statistical tables or formulas, and then dimensional deviation or uncertainty determined from Eqs. (13 & 14).

2.6. Data Outliers

If during repetitive measurements under apparently the same conditions, one or small number of measurements happen to be very much different from many other similar measurements, i.e. if the probability for such happening is very small, then such measurements, also called *outliers*, may be discarded. However, discarding "bad" measurements should not be a rule but rather a justified exception. Usually, if a measurement is outside $\pm 3S_x$ (i.e., $t\approx z=\pm 3$), it should be rejected since the probability for that to happen is extremely small, i.e., $P_{out}=1-0.9973=0.27\%<½\%$ (according to the Gaussian distribution for large number of measurements). For large number of measurements the single-tailed outside probability may be relaxed to $½P_{out}=0.1/n$, and the corresponding outlier z_{OL} (or t_{OL}) determined from the Gaussian (or the Student's *t*-) distribution. If a measurement's $|z| > z_{OL}$ (or $|t| > t_{OL}$) then it may be discarded (it is an *outlier*), otherwise it should be included in statistical analysis. For example, for $n=10$ measurements, $½P_{out}=0.1/n=0.01=1\%$, corresponds to $z=2.33$, thus all measurements outside $\pm 2.33S_x$ deviations should be discarded. Again, all effort should be taken to avoid having data outliers.

2.7. Relative Uncertainty

For any uncertainty (say u_{any}) the corresponding relative uncertainty ($u_{any,r}$) may be evaluated as a dimensionless ratio or percentage of the magnitude of measured quantity (e.g., x_{any}), i.e.:

$$u_{any,r} = \frac{u_{any}}{|x_{any}|} = \frac{u_{any}}{|x_{any}|}100\% \tag{17}$$

It is recommended and often advantageous to evaluate the corresponding relative uncertainty, and particularly the extended (resultant) relative uncertainty, in order to evaluate its relative fraction or percentage of the measured quantity. The uncertainty of a well-conducted measurement, by its very nature, is supposed to be small, a couple or several percent. Negligibly small percentage may indicate that not all critical sources of errors were accounted for, and too large percentage may indicate that an inaccurate instrument or inappropriate measurement methods were employed. Any extreme results should be reviewed and appropriate actions taken.

3. Putting It All Together: An Example

[NOTE: This example will be expended with more complexity and quantitative results in the future revisions of this manuscript, with an intention to include most of the concepts (and formulas) from this article].

If we are to measure power *P*, dissipated in a resistor *R* under supplied voltage potential difference *V*, the corresponding correlations will be:

$$P = P(V, R) = V^2 / R \tag{18}$$

$$u_P = \sqrt{(u_{P,V})^2 + (u_{P,R})^2} = \sqrt{\left(\frac{\partial P}{\partial V}u_V\right)^2 + \left(\frac{\partial P}{\partial R}u_R\right)^2} = \sqrt{\left(\frac{2V}{R}u_V\right)^2 + \left(\frac{V^2}{R^2}u_R\right)^2} \qquad (19)$$

However, if resistance is a known function of temperature T, say $R_T = R_0[1+b(T-T_0)]$, and the temperature T and voltage V are measured by an experimenter, as well as an additional measurement bias, B, is empirically estimated, then the power measurement equation (which accounts for all major contributing components) becomes:

$$P = P(V, R_0, b, T, T_0, B) = \frac{V^2}{R_0[1 + b(T - T_0)]} + B \qquad (20)$$

The uncertainty of the power is then determined using sensitivity and RSS Eqs. (11 & 12), as well as known or estimated uncertainties of all components ($u_V, u_{Ro}, u_b, u_T, u_{To}, u_B$), i.e.:

$$\begin{aligned} u_P &= \sqrt{(u_{P,V})^2 + (u_{P,Ro})^2 + (u_{P,b})^2 + (u_{P,T})^2 + (u_{P,To})^2 + (u_{P,B})^2} \\ &= \sqrt{\left(\frac{\partial P}{\partial V}u_V\right)^2 + \left(\frac{\partial P}{\partial R_0}u_{R_0}\right)^2 + \left(\frac{\partial P}{\partial b}u_b\right)^2 + \left(\frac{\partial P}{\partial T}u_T\right)^2 + \left(\frac{\partial P}{\partial T_0}\underbrace{u_{T_0}}_{\approx 0}\right)^2 + \left(\frac{\partial P}{\partial B}u_B\right)^2} \end{aligned} \qquad (21)$$

Note that B may have any value, including zero, still its uncertainty may exist, and the corresponding sensitivity $\partial P/\partial B=1$. Uncertainties much smaller that the others may be neglected (like $u_{To}\approx 0$). Details and evaluation of uncertainty components of a resultant uncertainties may differ, but if all critical sources of errors are properly accounted for, then the final results should be close to each other when expended to the same confidence level.

4. Conclusion

As already stated, the uncertainty quantification is very judgmental and based on full physical understanding of all related issues and accumulated previous experience (it is more an art than a science), however, the uncertainty analysis procedure is rather simple, and consists of the following steps:

(1) identification of all critical (i.e. the most contributing) uncertainty components (this is very important since the final result depends on what is accounted for);
(2) quantification of all standard uncertainties, based on relevant statistics, like the *Student's t-distribution* to account for small number of measurements (type A uncertainties), or obtained from other external sources (type B uncertainties);
(3) use the law of propagation of uncertainties (sensitivity coefficients based on the first-order Taylor series) when a partial or final measurement result is determined using an equation from other measured quantities;
(4) reduce all uncertainty components to standard uncertainties or to uncertainties for the same probability confidence level and combine them using the Root-Sum-of-Squares (RSS) rule;
(5) Expend the combined uncertainty to desired confidence level by multiplying it with a relevant coverage factor or dimensionless deviation *t*, based on the combined effective degree of freedom using the Student t-distribution.
(6) Finally, report all relevant details of accounted uncertainty components along with expended resultant uncertainty, but also some indication of the values of uncertainty components not accounted for, should always be reported.

It is important to clarify again, that only the uncertainty, i.e. a deviation range (or interval) from a reported measurement result, with corresponding probability, may be evaluated, but it is not possible to obtain a perfect (error-free) measurement, nor it is possible to estimate an uncertainty with 100% probability (absolute certainty is impossible too). However, under well-controlled conditions and well-understood measurement process and procedure, it is possible to minimize and reasonably well (at high probability) estimate the uncertainties of measured quantities and the final measurement result.

References

[1] Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results, NIST Technical Note 1297, 1994 Edition http://physics.nist.gov/Pubs/guidelines/TN1297/tn1297s.pdf , July 2003.

[2] Galleries of Probability Distributions: *NIST/SEMATECH e-Handbook of Statistical Methods*, http://www.itl.nist.gov/div898/handbook/eda/section3/eda366.htm , July 2003.

[3] R.H. Dieck, "Measurement Accuracy" a chapter in "The Measurement, Instrumentation and Sensors Handbook" *(J.G. Webster, Editor-in-Chief),* ISBN: 0-8493-8347-1, CRC Press, 1999.

[4] *NIST/SEMATECH e-Handbook of Statistical Methods*, http://www.itl.nist.gov/div898/handbook/, July 2003.

[5] Tables for Probability Distributions: *NIST/SEMATECH e-Handbook of Statistical Methods*, http://www.itl.nist.gov/div898/handbook/eda/section3/eda367.htm , July 2003.

[6] R.S. Figliola, and D.E. Beasley, *Theory and Design for Mechanical Measurements* - 3rd Edition, Wiley, 2000.